

Ulrich Schnörch

Der Autoren-Arbeitsplatz: Ein *elexiko*-Wortartikel entsteht

1. Der lexikografische Arbeitsplatz: Anforderungen, Arbeitsmittel und Methoden	105
2. Ein Wortartikel entsteht: <i>Verkehr</i>	112
2.1 Lesartenübergreifende Angaben	112
2.2 Lesartendisambiguierung	112
2.3 Lesartenbezogene Angaben	116
2.4 Ergänzende Informationen zu den lesartenübergreifenden bzw. lesartenbezogenen Angaben)	127
3. <i>elexikografie</i> – Ausblick	128
4. Literatur	130
4.1 Forschungsliteratur	130
4.2 Internetressourcen	130

1. Der lexikografische Arbeitsplatz: Anforderungen, Arbeitsmittel und Methoden

In der Einleitung zu diesem Band schreibt U. Haß:

Die Autoren erarbeiten ein Informationspotenzial, das unterschiedlichste Funktionen des einsprachigen Wörterbuchs und unterschiedliche Nutzungsinteressen unterschiedlicher Adressatengruppen abdecken kann. Erst durch die hypertextuelle Darstellung gegenüber den Nutzern werden je nach Situation und Interesse Informationen ausgewählt. (U. Haß, *elexiko* – Das Projekt, 2 f.)¹.

Diese Zeilen lassen leicht erahnen, dass die Erarbeitung des Informationspotentials für *elexiko*-Wortartikel in Bezug auf die anzuwendenden Verfahren zur Ermittlung der erforderlichen Datenquantität und -qualität ein äußerst komplexer Vorgang ist. Er kann in seiner gesamten Bandbreite und seinem Nuancenreichtum nicht auf wenigen Seiten nachvollzogen werden. Das ist streng genommen auch nicht nötig, denn die einzelnen Beiträge dieses Bandes widmen sich bereits detailliert den vielfältigen prinzipiellen, konzeptionellen und verfahrenstechnischen Schwerpunkten von *elexiko*, die LexikografInnen bei der Wortartikelproduktion unter anderem berücksichtigen müssen. Ausgehend von einem konkreten Beispiel wird die Entstehung eines *elexiko*-Wortartikels stark komprimiert und exemplarisch gezeigt. Der Hauptakzent liegt in der Zusammenfassung der Abfolge und des Zusammenspiels der einzelnen methodischen

¹ Die Beiträge aus dem vorliegenden Sammelband werden nachfolgend stets in der Form ‚Name, Kurztitel, ggf. Seitenangabe‘ zitiert. Damit sollten die Beiträge zweifelsfrei und leicht anhand des Inhaltsverzeichnisses zu identifizieren sein.

Ansätze (des sog. Workflows). Für die Diskussion der mit den einzelnen Arbeitsschritten einhergehenden theoretischen und praktischen Probleme und der entwickelten Lösungsstrategien wird auf die einschlägigen Artikel dieses Bandes verwiesen.

Der Spielraum für die Quantität und Qualität der von den LexikografInnen für einen Wortartikel mitunter zu verarbeitenden Daten kann sehr stark variieren. Mögliche Einflussfaktoren sind beispielsweise die Wortart (vgl. A. Klosa, Grammatik, in diesem Band, 282), die Anzahl der Korpusbelege² oder die Parametereinstellungen bei Korpusabfragen. Die Menge und das Spektrum der für einen Wortartikel zusammenzutragenden Informationen in Verbindung mit deren Auszeichnung im Rahmen der DTD (siehe C. Müller-Spitzer, Die Modellierung lexikografischer Daten, in diesem Band) sind die Voraussetzung, um diverse Varianten von Datenbankabfragen durchzuführen, und um mithilfe sog. Stylesheets spezifizierte Datenpräsentationen für unterschiedlichste Nutzungssituationen und Nutzergruppen zu entwickeln.

Folgendes kommt noch hinzu: Die Verfahren (s. u.), derer sich die LexikografInnen zur Ermittlung der lexikografisch relevanten Daten bedienen sind z. T. eigens für *lexiko* entwickelt bzw. angepasst worden: Sie haben den Elchtest im Zuge der Bearbeitung des Demonstrationswortschatzes³ bestanden. Dabei ist jedoch zu anmerken, dass die gewonnen Erfahrungen im praktischen Umgang mit den innovativen Methoden und Instrumenten sowie deren Verfeinerung und Abstimmung aufeinander Ergebnis einer relativ kurzen Zeitspanne sind.

Mit Hilfe eines Vergleichs aus dem Bereich des späteren Beispiels *Verkehr* lässt sich also zusammenfassen:

Der im Anschluss skizzierte Weg vom *lexiko*-Stichwort zum *lexiko*-Wortartikel führt vorwiegend über Hauptstraßen, die weitestgehend ohne Staus, Umleitungen, Sackgassen, Sichverfahren etc. ans Ziel führen – im lexi-

² Die Größe des *lexiko*-Korpus kann bei bestimmten Stichwörtern zu sehr hohen Trefferzahlen führen, daher wurde die Obergrenze in solchen Fällen per Zufallsauswahl auf 10000 festgelegt. Wenn in derartigen Fällen die Ergebnisse der Kookkurrenzanalysen als solche bei mehreren hintereinander durchgeführten Analysen weitestgehend deckungsgleich bleiben, basieren sie nicht auf der jeweils identischen Belegauswahl. Auch sehr niedrige Trefferzahlen reduzieren in rein statistischer Hinsicht natürlich die Aussagekraft der Korpusdaten.

³ Der Demonstrationswortschatz enthält 240 Stichwörter, die die Bandbreite und Strukturierung der Informationen illustrieren. Diese Wortartikelmenge soll die systematischen Möglichkeiten der Artikelstruktur bzw. des Informationsprofils von *lexiko* möglichst vollständig abdecken (z. Z. fehlen noch Angaben zur Wortbildungsproduktivität und zur Diachronie). Bei der Auswahl dieser 240 Stichwörter kamen zwei Kriterien zur Anwendung: Ausgewählt wurden erstens die mittels der COSMAS-Kollokationsanalyse ermittelten signifikanten Kookkurrenzpartner des (willkürlich festgelegten) Nomens *Mobilität*; zweitens wurde diese Menge systematisch ergänzt nach den aus Wörterbüchern bekannten Proportionen von Wortarten, Alphabetstrecken und diversen Wortbildungstypen; weitere Informationen dazu unter <http://www.lexiko.de/>.

kografischen (wie im alltäglichen) Leben freilich eine Idealvorstellung. Alle realiter zu bewältigenden Aufgabenstellungen müssen zwangsläufig vernachlässigt werden zu Gunsten eines stark vereinfachenden Werkstattberichts; der Schwerpunkt liegt dabei auf dem verfahrenstechnischen Prozess, der Abfolge der Arbeitsschritte ausgehend von einem ganz konkreten Beispiel.

Wie bereits angedeutet basieren die Erfahrungen mit der Wortartikelarbeit auf der lexikografischen Erschließung des Demonstrationswortschatzes (siehe Anm. 2) durch die LexikografInnen. Folglich steht primär die (Vorgehensweise zur) Datenermittlung im Vordergrund, d. h. es geht hier um das, was die LexikografInnen in das System einspeisen und wie sie es ermitteln. Es geht hier nicht um realisierbare Präsentationsmöglichkeiten der *lexiko*-Wortartikel, die je nach Benutzungsgruppe und Benutzungssituation sehr unterschiedlich ausfallen können. Grund und Grundlage hierfür ist zunächst die Art der Datenmodellierung⁴; sie ermöglicht die Entwicklung geeigneter Stylesheets, mit deren Hilfe wiederum verschiedene Online-Ansichten für *lexiko*-Wortartikel generiert werden können – je nach spezifizierter Nutzungssituation und NutzerInnengruppe. An dieser nutzungsorientierten Präsentation der Daten wird in Zukunft noch stärker zu arbeiten sein.

Bei *lexiko* handelt es sich prinzipiell um ein dynamisches System, das fortlaufend Änderungen, Verbesserungen, Verfeinerungen etc. unterworfen ist, an denen die NutzerInnen kontinuierlich teilhaben können – anders als im gedruckten Wörterbuch, in dem Neuerungen gebündelt erst in einer Folgeauflage berücksichtigt werden können. *lexiko* funktioniert einmal wie ein Wörterbuch, d. h. man kann vergleichbar dem gedruckten Nachschlagwerk in gewohnter Weise recherchieren; zusätzlich wird *lexiko* auch verstärkt als Informationssystem zu nutzen sein, um Anfragen über den gesamten (bearbeiteten) Wortschatz hinweg ausführen zu können. Daher muss ein sehr breites Spektrum an Informationen durch die LexikografInnen vorab erarbeitet und eingetragen werden (von denen – um es nochmals zu betonen, anders als im gedruckten Wörterbuch – nicht alle zugleich angezeigt werden müssen).

Auf den Punkt gebracht: Im Mittelpunkt steht nachfolgend explizit die Perspektive der LexikografIn, also die produzierende Seite – die rezipierende wird insofern implizit berücksichtigt, als sie die Arbeit der LexikografInnen maßgeblich beeinflusst, indem sie Antworten auf mögliche Fragen potenzieller NutzerInnen in unterschiedlichsten Situationen zu geben versuchen.

4

„In einem Projekt wie *lexiko* ist daher eine [...] möglichst granulare Kennzeichnung der einzelnen Angaben – unabhängig von ihrer konkreten Darstellung – von zentraler Wichtigkeit. Ziel ist die Erstellung einer Datenbasis, aus der mehrere Präsentationen entwickelt werden können“ (C. Müller-Spitzer, Die Modellierung lexikografischer Daten, in diesem Band, S. 42).

Eine Skizze des allgemeinen Rahmens, in den die Arbeit des *elexiko*-LexikografInnen eingebettet ist, wäre unvollständig ohne die Zusammenstellung einiger wichtiger Grundlagen und Arbeitsmittel:

- *elexiko*-Korpus (siehe dazu: P. Storjohann, *elexiko*-Korpus, in diesem Band); mit ihm wird das grundlegende Prinzip der Korpusbasiertheit in *elexiko* umgesetzt; die lexikografische Beschreibung auf der Grundlage eines durch Texte repräsentierten Sprachausschnitts setzt natürlich einen gewissen Grad an Usualität der beschriebenen Phänomene voraus und ist deskriptiv ausgerichtet (siehe dazu U. Haß, *elexiko* – Das Projekt, in diesem Band, und A. Klosa, Sprachkritik und Sprachreflexion, in diesem Band).
- COSMAS II⁵ und das integrierte Tool ‚Statistische Kollokationsanalyse und Clustering‘ dienen v. a. der Vorstrukturierung, Analyse und Sortierung von Massendaten, der Belegauswahl usw., und liefern somit zentrale Antworten auf Korpusanfragen.
- XML-Editor: X-MetaL (siehe dazu C. Müller-Spitzer, Modellierung, in diesem Band, 45, bes. Anm. 8).
- D(ocument) T(ype) D(efinition) sowie deren Dokumentation (siehe dazu C. Müller-Spitzer, Modellierung, in diesem Band). Es ist geplant, in die DTD-Dokumentation später auch redaktionelle Leitlinien und Absprachen zu integrieren, wie sie traditionellen Redaktionshandbüchern zu Grunde liegen.
- objekt-relationale Datenbank (Oracle), mit deren Hilfe u. a. die Stichwortliste und die Wortartikel verwaltet werden (siehe dazu C. Müller-Spitzer, Modellierung, in diesem Band).
- Sekundärliteratur, v. a. Wörterbücher (siehe z. B. die Zusammenstellung der Referenzwerke von A. Klosa, Sprachkritik und Sprachreflexion, in diesem Band, 303 f.).

Im Umfeld der Stichwortbearbeitung ist es zudem wichtig, grundsätzlich eine Reihe von Begriffspaaren zu klären, die für *elexiko* konzeptionell von Bedeutung sind, und die den Prozess der Artikelproduktion stark beeinflussen:

- **lexikografisch** vs. **automatisch** erzeugte Daten: Hierbei handelt es sich einerseits um Daten für Wortartikel, welche automatisch, d. h. mittels Computer und Software erzeugt werden, sowie um Daten, die die LexikografInnen (z. T. auf der Vorarbeit des Computers) intellektuell ableiten und erarbeiten. Aus Gründen der Transparenz sind

⁵ Siehe <http://www.ids-mannheim.de/cosmas2/>

diese beiden Vorgänge methodisch strikt zu trennen, auch wenn sie häufig ineinander greifen.⁶

- **corpus-driven (CD) vs. corpus-based (CB) Analyse:** Hierbei handelt es sich um zwei Methoden der Korpusbefragung, einmal der unvoreingenommenen mittels statistischer Abfrageroutinen (CD), zum anderen der hypothesengeleiteten (siehe dazu P. Storjohann, Paradigmatische Relationen, in diesem Band, 251 ff.)
- **lesartenübergreifende vs. lesartenspezifische Angaben:** Hierbei handelt es sich um Angaben, die zum Stichwort als Ganzes oder zu seinen einzelnen Lesarten bzw. Lesartenspezifizierungen gemacht werden können bzw. müssen.
- **obligatorische vs. fakultative Angaben:** Hierbei handelt es sich um Angaben, die bei der Artikelproduktion ermittelt und ausgefüllt werden müssen und um solche, die ausgefüllt werden können (mögliche Zwischenstufen werden hier vernachlässigt); dies kann einzelne Angaben betreffen (siehe C. Müller-Spitzer, Modellierung, in diesem Band, 56 ff.), aber auch ganze Bereiche wie z. B. die Wortbildungsproduktivität, die (aus Zeitgründen) bislang nicht in den Wortartikeln behandelt wurde (siehe A. Klosa, Wortbildung, in diesem Band)

Aufgabe der Lexikografinnen ist es nun, in einer geregelten Abfolge von Arbeitsschritten mit Hilfe der aufgelisteten Grundlagen, Instrumente und Methoden lexikografische Daten für lesartenübergreifende und lesartenspezifische, obligatorische (und ggf. fakultativen) Angaben eines Wortartikels zu ermitteln und von der DTD unterstützt in die XML-Instanz einzutragen.

Für die Erarbeitung der *lexiko*-Wortartikel ist die herausragende Bedeutung eines Prinzips nochmals hervorzuheben: das Prinzip der **Korpusbasiertheit**. Darunter ist zu verstehen, dass das Gros der Informationen in den *lexiko*-Wortartikeln auf Befragung und Auswertung einer riesigen digitalen Textsammlung, des *lexiko*-Korpus (siehe P. Storjohann, *lexiko*-Korpus, in diesem Band) basiert. Dies kann nur geleistet werden durch Zuhilfenahme eines geeigneten Rechercheinstruments, wodurch es erst möglich ist, die Massendaten nach Bedarf vorzustrukturieren, zu analysieren und zu sortieren; um einen *lexiko*-Wortartikel schreiben zu können, müssen Lexikografinnen entsprechende Recherchen mit COSMAS II durchführen sowie das Tool ‚Kollokationsanalyse‘ nutzen, und die jeweiligen Anfrageergebnisse zu interpretieren

6

Die Ergebnisse der automatischen Kollokationsanalyse (und deren Analyse) bilden die Datenbasis für die Lesartendisambiguierung im Rahmen der theoretischen Klassifikation, für die Auswahl der semantischen Mitspieler, für die Ableitung der typischen Verwendungen und der sinnverwandten Wörter bis hin zur Formulierung der Phrasenmuster zur Valenzangabe.

wissen. Die CD-Methode gibt dabei die Richtung vor, u. U. gefolgt von weiteren CB-Abfragen zur hypothesengeleiteten, spezifizierenden Korpusbefragung.

Auf diesem Gebiet wurde und wird für die deutschsprachige Lexikografie viel Pionierarbeit geleistet und trotz aller konzeptioneller Vorüberlegungen ist dem Artikelschreiben in diesem Zusammenhang ein fortgeschrittener Experimentierstatus kaum abzusprechen. Entsprechend kann es vorkommen, dass Abläufe, wie sie momentan noch stattfinden, mit zunehmendem Erkenntnisgewinn geändert werden. Die empirisch gestützte, fortlaufende Optimierung des sog. Workflows liegt auf einer Linie mit dem dynamischen Prinzip eines Online-„Wörterbuchs“: Das elektronische Medium verkürzt die Reaktionszeiten auf Veränderungen (im Gegensatz zum gedruckten) erheblich.

Die ermittelten Daten gilt es schließlich noch festzuhalten, d. h. den Wortartikel im eigentlichen Sinne zu schreiben. Dies geschieht mithilfe eines XML-Editors (im Falle von *exiko*: XMetaL). Für das Abfassen der *exiko*-Wortartikel wurde eigens eine komplexe DTD entwickelt (siehe C. Müller-Spitzer, Modellierung, in diesem Band): Sie erlaubt es den LexikografInnen die Fülle der gewonnenen Daten im Rahmen einer vorgegebenen Struktur einzutragen und auszuzeichnen. Die DTD und deren Dokumentation als maßgebliche „Unterstützung der Lexikografen in der formalen Einhaltung der Artikelstruktur“ (C. Müller-Spitzer, Modellierung, in diesem Band 4, siehe dort auch zu weiteren Einzelheiten, Bearbeitungskonventionen etc.), wird noch ergänzt von einem Redaktionshandbuch im traditionellen Sinne. Darin wurden redaktionelle Absprachen (wie Richtlinien für bestimmte Formulierungen, Formatierungen, Markierungen etc.) festgehalten, die im Zuge der lexikografischen Erarbeitung des Demonstrationswortschatzes getroffen wurden. Das Redaktionshandbuch soll später in XML überführt und in die DTD-Dokumentation integriert werden.

Die redaktionell bearbeiteten Wortartikel⁷ werden schließlich – sofern sie valide sind, d. h. den Regeln der DTD entsprechen (vgl. C. Müller-Spitzer, Modellierung, in diesem Band, 56 ff.) – in der objekt-relationalen Datenbank Oracle abgelegt, gespeichert (vgl. ebd., 44) und in dieser verwaltet. In der Datenbank befinden sich bereits alle knapp 300.000 Stichwörter der *exiko*-Stichwortliste (siehe U. Schnörch, *exiko*-Stichwortliste, in diesem Band) als angelegte Minimalartikel (s. u.). Aus diesen lässt sich modular nach bestimmten Kriterien eine mehr oder minder große Anzahl auswählen, die lexikografisch zu bearbeiten ist; einer solchen Menge entspricht auch der Demonstrationswortschatz, der zu Testzwecken zusammengestellt und bearbeitet wurde.

7

Im Online-Wörterbuch scheut man den Ausdruck ‚fertig‘, weil jeder Artikel prinzipiell beliebig oft überarbeitet, verändert z. B. aktualisiert, und wieder ‚veröffentlicht‘ werden kann

2. Ein Wortartikel entsteht: *Verkehr*

2.1 Lesartenübergreifende Angaben

Einen Teil der lesartenübergreifenden Angaben finden die Lexikografinnen bereits vor, denn darunter fallen die eben erwähnten (teil-) automatisch gewonnenen Angaben zum Stichwort (vgl. Abbildung 2).

Verkehr

Orthografie

Normgerechte Schreibung *Verkehr*

Worttrennung *Ver|kehr*

Abbildung 2: Orthografische Angabe im Wortartikel *Verkehr*

Informationen zur Wortbildung sind hingegen von den Lexikografinnen zu eruieren und einzutragen¹⁰; im Falle von *Verkehr* kann keine Angabe gemacht werden, d. h. das Stichwort kann synchron unter dem Aspekt der Wortbildung nicht näher beschrieben werden.

An der Gelenkstelle zwischen den lesartenübergreifenden und den lesartenbezogenen Angaben müssen die Lexikografinnen das Bedeutungsspektrum des Stichworts mit dem Ziel analysieren, dessen Lesart(en) und ggf. die Spezifizierung(en) bestimmen (siehe dazu U. Haß, Das Bedeutungsspektrum, in diesem Band).

2.2 Lesartendisambiguierung

Für die Analyse des Bedeutungsspektrums ist in *elexiko* das Prinzip der Korpusbasiertheit maßgebend, ergo müssen die Lexikografinnen an dieser Stelle mit der Befragung des *elexiko*-Korpus beginnen. Zu diesem Zweck wurde in COSMAS II eine Abfrage für das Lemma *Verkehr* durchgeführt: Die insgesamt knapp 87000 Treffer¹¹ wurden durch Zufallsauswahl auf 10000 Treffer reduziert; sie bilden die Basis für die anschließende Kollokationsanalyse, die

¹⁰ Entgegen erster Annahmen hatte sich im Verlauf der Arbeiten zum Demonstrationswortschatz gezeigt, dass Angaben zur Wortbildung nicht in jedem Fall auf der Stichwortebene anzusiedeln sind; es gibt Fälle, in denen dies auf der Lesartenebene zu erfolgen hat (siehe den Wortartikel *bereiten* unter <http://www.elexiko.de/>). Die entsprechende Datenmodellierung wird in absehbarer Zeit geändert (siehe A. Klosa, Wortbildung, in diesem Band).

¹¹ Sie verteilen sich auf die Wortformen: VERKEHR: 1.389, Verkehr: 72.741, Verkehrs: 12.455, verkehrs: 322, VERKEHRS: 5.

als Ergebnis 336 Kookkurrenzpartner (und die dazu gehörigen syntagmatischen Muster) erbrachte – daneben 3817 statistisch unspezifische Fälle. Zu Illustrationszwecken genügt ein kleinerer Ausschnitt aus dem mittleren Bereich der gesamten Kookkurrenzliste (sie ist nach absteigender Signifikanz der Kookkurrenzpartner von *Verkehr* geordnet und zeigt die [typische] syntagmatische Einbettung der Kollokation; sie wurde der besseren Übersichtlichkeit halber um einige numerische Parameter gekürzt):

Bau	Bau und der ... <i>Verkehr</i>
Strassen	Strassen mit ... <i>Verkehr</i>
anziehen viel	viel <i>Verkehr</i> anziehen
anziehen	<i>Verkehr</i> [...] anziehen
motorisierte	motorisierte <i>Verkehr</i>
Arbeitskreis	Der Arbeitskreis [...] <i>Verkehr</i>
Verschönerungsvereins	des <i>Verkehrs</i> und Verschönerungsvereins
Autobahn	<i>Verkehr</i> auf der Autobahn
Wissenschaft Bundesminister	Bundesminister für Wissenschaft und <i>Verkehr</i>
Wissenschaft	Wissenschaft [und] <i>Verkehr</i> und ...
Mannheimer	die der Mannheimer [...] <i>Verkehrs</i> AG MVG
geschäftlichen	im geschäftlichen <i>Verkehr</i>
Finanzierung	<i>Verkehr</i> ... Finanzierung
Behinderungen	zu Behinderungen des <i>Verkehrs</i>
dichter	dichter <i>Verkehr</i>
geöffnet	für den <i>Verkehr</i> geöffnet
Sektion Wirtschaftskammer	Obmann von der Sektion <i>Verkehr</i> in der Wirtschaftskammer in ...
Sektion	Sektion <i>Verkehr</i>
fließende	fließende <i>Verkehr</i>
Nachrichtenübermittlung Bereich	Bereich <i>Verkehr</i> und Nachrichtenübermittlung
Nachrichtenübermittlung	<i>Verkehr</i> [und] Nachrichtenübermittlung
sperrern	für den <i>Verkehr</i> [zu] sperren
Telekommunikation	<i>Verkehr</i> und Telekommunikation
Themen Soziales	Themen ... Soziales ... <i>Verkehr</i>
Themen	zu den die Themen [...] und] <i>Verkehr</i> und
Sachen	in Sachen <i>Verkehr</i>
Landesamt Straßenbau	Landesamt für <i>Verkehr</i> und Straßenbau
Landesamt	das ... Landesamt für <i>Verkehr</i>
Bereich	im Bereich [...] <i>Verkehr</i>
Soziales Sicherheit Finanzen	Finanzen ... <i>Verkehr</i> Sicherheit ... Soziales
Soziales Sicherheit	Sicherheit Soziales ... <i>Verkehr</i>
Soziales	Soziales [...] <i>Verkehr</i>
Strasse	<i>Verkehr</i> auf von der ... Strasse
fließt	dass Der der <i>Verkehr</i> [...] fließt über
Versorgungsbetriebe	<i>Verkehrs</i> und Versorgungsbetriebe
viel	viel [...] <i>Verkehr</i>
Landwirtschaft	Landwirtschaft und ... <i>Verkehr</i>

Sicherheit
internationalen
zeitweise
rollen
Landesausstellung

stockendem
worden
zusätzlicher
Industrie Gewerbe
Industrie
Straßenwesen
Bundesamtes
Infrastruktur
Waren freien
Waren
rollte
Richtungen

legten
Bundesminister
innerdeutschen
überwachen
fließen
Mehr

teilweise
gebracht
innerstädtischen
Tarifverbunds

freien
stockender
Staus
ungeschützten
still
flüssigeren
Erschliessung
Umweltplanung IVU

die Sicherheit [im und ...] *Verkehr*
im|den|Im internationalen [...] *Verkehr*
Verkehr auf ... zeitweise
der *Verkehr* [...] wieder [...] rollen
die Landesausstellung [...] *Verkehr* in
Knittelfeld
stockendem *Verkehr*
für im *Verkehr* ... worden
zusätzlicher *Verkehr*
Verkehr Gewerbe ... Industrie
aus Industrie [und] *Verkehr* und
Verkehr ... Straßenwesen
des Bundesamtes für *Verkehr*
die ... Infrastruktur [für] *Verkehr*
freien *Verkehr* ... Waren
Verkehr ... Waren
rollte der *Verkehr*
der|den *Verkehr* aus|in beide|beiden
Richtungen
legten den *Verkehr*
Bundesminister ... *Verkehr*
im innerdeutschen [...] *Verkehr*
den *Verkehr* [...] überwachen
der *Verkehr* [...] wieder] fließen
IG ... Mehr [Leben und weniger]
Verkehr für alle
der *Verkehr* [...] teilweise
in den *Verkehr* gebracht
innerstädtischen *Verkehr*
des Frankfurter *Verkehrs* und
Tarifverbunds FVV ...
den freien [...] *Verkehr* von der
stockender *Verkehr*
Verkehr und Staus zu ...
ungeschützten *Verkehr*
stand|steht der *Verkehr* [in stand] still
flüssigeren *Verkehr*
Verkehr ... Erschliessung
100% für Informatik *Verkehrs* und
Umweltplanung [...] IVU

Die gesamte Auflistung als Ergebnis der corpus-driven Analyse von Massendaten (s. o.) ist ein wichtiges lexikografisches Hilfsmittel für die folgenden Arbeitsschritte (s. u.); obgleich sich die LexikografInnen darüber hinaus, und unter Umständen von dieser Vorstrukturierung angeregt weiterer Analysemethoden bedienen müssen (mittels des corpus-based Verfahrens zur Stützung von Hypothesen, die ggf. aus der näheren Betrachtung der statistisch unspezifischen Fälle herrühren), steht die Auswertung des Kookkurrenzliste im Mittelpunkt, bildet gleichsam den roten Faden. Natürlich ist es von Vorteil, dass das Ergebnis der Kookkurrenzanalyse variabel darstellbar ist: Zu den einzelnen Kookkurrenzpartnern (aber auch für alle statistisch unspezifischen Fälle) kön-

nen beispielsweise je nach Bedarf KWICs (= Key Word in Context) erzeugt werden, also Listen von unterschiedlich sortierbaren Zeilen, die das abgefragte Wort jeweils in der unmittelbaren Satzumgebung zeigen. Hierzu können die kompletten, im Kontextumfang wählbaren Belegstellen aufgerufen werden usw.¹²

Entsprechend der Richtlinien zur Lesartendisambiguierung (siehe U. Haß, Das Bedeutungsspektrum, in diesem Band) und dem Befund der Kollokationsanalyse wurden für *Verkehr* zwei Lesarten festgesetzt und in der für *lexiko* gemäßen Weise etikettiert und in Abhängigkeit zur satzfunktionalen Klasse nach musterhaftem Vorbild paraphrasiert (siehe hierzu P. Storjohann, Semantische Paraphrasen und Kurzetikettierungen, in diesem Band). Eine Lesart drängt sich aufgrund der überwiegenden Mehrzahl der Kookkurrenzpartner förmlich auf, die Lesart ‚Fortbewegen‘ (vgl. Abbildung 3).¹³

Verkehr Lesart ‚Fortbewegen‘

Bedeutungserläuterung

Mit **Verkehr** bezieht man sich auf die Gesamtheit der Menschen und Güter, die auf der Straße, der Schiene, dem Wasser und in der Luft mittels geeigneter Fortbewegungsmittel von einem Ort zum anderen gelangen bzw. transportiert werden. In metaphorischer Weise kann mit **Verkehr** auch auf die Weiterleitung von Daten Bezug genommen werden.

Abbildung 3: Bedeutungserläuterung zum Stichwort *Verkehr*. Lesart ‚Fortbewegen‘

Auf eine zweite Lesart deutet im ausgewählten Abschnitt zumindest ein Partner hin, nämlich *ungeschützten*. Diesem ersten Hinweis folgt die gezielte Untersuchung des Gesamtbefundes (also auch der statistisch unspezifischen Partner sowie der Gesamtrefferliste), woraufhin in der Tat eine weitere usuelle Lesart angesetzt werden kann, nämlich: die Lesart: ‚Beischlaf‘ (vgl. Abbildung 4).

¹² Zu Art und Funktion der einzelnen Ergebnispräsentationen von Korpusabfragen siehe U. Haß, *lexiko* – Das Projekt, in diesem Band.

¹³ Die einzelnen Lesarten werden in *lexiko* nicht wie traditionell in Printwörterbüchern mit numerischen Indizes, sondern mit Kurzetiketten versehen. Hierbei handelt es sich soweit als möglich um Synonyme bzw. Hyperonyme, die einzig die Funktion der Identifizierung von Lesarten und deren Spezifizierungen besitzen (nähere Einzelheiten siehe P. Storjohann, Semantische Paraphrasen und Kurzetikettierungen, in diesem Band).

Verkehr Lesart ‚Beischlaf‘

Bedeutungserläuterung

Mit **Verkehr** bezeichnen Sprecher die Handlung, bei der zwei Personen sexuellen Kontakt (meist Beischlaf) miteinander haben.

Abbildung 4: Bedeutungserläuterung im Stichwort *Verkehr*, Lesart ‚Beischlaf‘

Während es sich bei der Lesart ‚Fortbewegen‘ um ein Kollektivum handelt kann die Lesart ‚Beischlaf‘ den durativen Handlungsprädikatoren zugeordnet werden. Mit der Zuordnung zu unterschiedlichen Prädikatorenklassen wird zugleich ein entscheidendes Kriterium für das Ansetzen zweier Lesarten erfüllt (siehe U. Haß, Das Bedeutungsspektrum, in diesem Band). Ein durch Konzeptverschiebung erklärbarer Zusammenhang der beiden Lesarten ist nicht zu erkennen, sodass die betreffende Angabe nicht ausgefüllt werden kann bzw. muss (siehe hierzu: U. Haß, Das Bedeutungsspektrum, in diesem Band).

2.3 Lesartenbezogene Angaben

Das lexikografisch fassbare Ergebnis der Disambiguierung des Stichworts *Verkehr* (und eigentlich bereits vorweggenommene lesartenbezogene Angaben) sind zwei etikettierte und semantisch paraphrasierte Lesarten, die mit (maximal) drei Korpusbelegen illustriert sind (zu Belegen in *lexiko* siehe A. Klosa, Belege, in diesem Band, zu ersten Vorüberlegungen zu Illustrationen siehe C. Müller-Spitzer, Modellierung, in diesem Band). Die Belegauswahl erfolgte praktischerweise in einem Zug mit der Analyse des Bedeutungsspektrums; sie kann durch die Lexikografinnen auch gesteuert werden, da per Mausklick schnell zwischen den Ansichten ‚Kookkurrenzergebnisse‘, ‚KWIC‘ und eben ‚Volltext‘ (= im Umfang variierbare Belegstellen) gewechselt werden kann. Erst auf Grundlage der Bedeutungsanalyse und der daraus resultierenden Lesart(en)festlegung können die Lexikografinnen jetzt der Reihe nach die lesartenbezogenen Angaben für das Stichwort *Verkehr* erarbeiten, die sich unter fünf Bereiche subsumieren lassen:

- semantische Umgebung und lexikalischen Mitspieler,
- typische Verwendungen,
- sinnverwandte Wörter,
- Besonderheiten des Gebrauchs,
- Grammatik.

Die Ermittlung der entsprechenden Angaben fußt gleichfalls auf dem Prinzip der Korpusbasiertheit. Für dessen methodische Umsetzung ist – wie oben angedeutet – die Kollokationsanalyse von zentraler Bedeutung: Lexikografinnen greifen im Zuge der folgenden Arbeitsschritte immer wieder auf die bereits zur Lesartendisambiguierung herangezogene Kookkurrenzliste zurück. Je nach Intention lässt sich diese modifizieren (z. B. Änderung der Parametereinstellungen, etwa des einbezogenen Analysekontexts) und in der Ansicht variieren (s. o.). Darüber hinaus kann es auch nötig werden, einzelnen aus dieser Analyse gewonnenen Vermutungen oder Hypothesen mit neuerlichen Anfragen ans Korpus nachspüren. Wie die Lexikografinnen im Prozess einer fortlaufenden Korpusbefragung die Abfragetools und -parameter je nach Erkenntnisinteresse im Detail anwenden, aufeinander abstimmen und rückkoppeln, kann von Stichwort zu Stichwort, von Analyseergebnis zu Analyseergebnis abweichen, im Großen und Ganzen sind Richtung und Spielräume jedoch vorgegeben. Die folgenden Teilkapitel sollen genau davon ein Bild vermitteln.

2.3.1 Semantische Umgebung und lexikalische Mitspieler

Welcher Art die Informationen sind, die Lexikografinnen unter dieser Überschrift erarbeiten, ist in der Online-Ansicht an der betreffenden Stelle in einer aufklappbaren Infobox dokumentiert (siehe <http://www.elexiko.de/>):

Die meisten Wörter verlangen aus Gründen ihrer Bedeutung bestimmte Arten anderer Wörter in ihrem engeren Textzusammenhang; der gewählte Satzbau spielt hierbei eine untergeordnete Rolle. Jedes Wort steht unabhängig davon, welches Satzglied es darstellt und wie es syntaktisch verknüpft ist, im Zentrum eines semantischen Netzes aus Mitspieler-Wörtern. Die Fäden, die das Stichwort in seiner jeweiligen Lesart mit seinen Mitspieler-Wörtern verbinden, sind unterschiedlicher Art. Diese Art wird in *elexiko* in Form einer einfach verständlichen Frage ausgedrückt, zu denen die Mitspieler-Wörter die Antworten liefern. Alle Mitspieler-Wörter sind aus dem *elexiko*-Korpus gewonnen; sie kommen häufig und typischerweise mit dem Stichwort im selben Textzusammenhang vor. Sie werden hier in ihrer jeweiligen Normalform (z. B. Nominativ Singular, Infinitiv) angegeben.

Diese Form der Angabe in *elexiko* deckt ab, was in der Linguistik häufig Argumentstruktur und semantische Rollen genannt wird. An die Stelle abstrakter Rollenkategorien oder sog. slots treten natürlichsprachliche Fragen; auch die sog. filler werden nicht typisiert, sondern als Reihe signifikanter Lexemrealisierungen repräsentiert. (<http://www.elexiko.de/>; zu Einzelheiten siehe U. Haß, Semantische Umgebung, in diesem Band)

Ausschlaggebend für Anzahl und Formulierung der Fragen, als deren Antworten die lexikalischen Mitspieler aufgezählt werden, sind zwei Dinge: einmal Fragesets, die zur Groborientierung musterhaft für alle angesetzten Prädikatenklassen in der Sammlung fürs Redaktionshandbuch zusammengestellt

wurden; zum anderen die Untersuchung der aus der Kollokationsanalyse gewonnenen Kollokatoren¹⁴ mit dem Ziel der semantischen oder diskursiv-thematischen Gruppierung. Diese ist letztlich ausschlaggebend für die tatsächliche Zahl und konkrete Formulierung der Fragen zur Mitspielerangabe für die Lesart eines Stichworts.

Im oben abgedruckten Ausschnitt der Kookkurrenzanalyse zu *Verkehr* findet man folgende Zeilen:

geschäftlichen
motorisierte
zusätzlicher

im geschäftlichen *Verkehr*
motorisierte *Verkehr*
zusätzlicher *Verkehr*

Wenn man die Gesamtliste zurate zieht, findet man dort zusätzlich u. a.:

öffentlichen
motorisierten
gesamten werden
[...]

für den öffentlichen [...] *Verkehr*
für den motorisierten *Verkehr*
gesamten *Verkehr* ... werden

Für den kollektiven Gattungsprädikatoren *Verkehr* in der Lesart ‚Fortbewegen‘ können die genannten Mitspieler als Antworten auf die Frage ‚Wie wird Verkehr modifiziert?‘ verstanden werden. Alle für den *elexiko*-Wortartikel *Verkehr*, Lesart ‚Fortbewegen‘ ermittelten Antworten (in alphabetischer Reihenfolge) auf die Frage *Wie wird Verkehr modifiziert?* lauten:

anwachsend, alltäglich, entgegenkommend, gesamt, geschäftlich, hausgemacht, landwirtschaftlich, motorisiert, nachfolgend, nichtmotorisiert, normalmotorisiert, öffentlich, privat, schienenengebunden, steigend, täglich, unnötig, viel, visafrei, wachsend, zunehmend, zusätzlich, zu viel.

Aus der Kookkurrenzliste für *Verkehr*/‚Fortbewegen‘ ließen sich des Weiteren Antworten(gruppen) auf folgende Fragen zusammenstellen (zu den einzelnen Mitspielern siehe den Artikel *Verkehr* unter <http://www.elexiko.de/>)

Wie wird die räumliche Aufteilung von Verkehr modifiziert?
Was tut Verkehr?
Was geschieht mit dem Verkehr?
Wo findet Verkehr statt?
Welche Verkehrsmittel sind am Verkehr beteiligt?
Was wird im Zusammenhang mit Verkehr thematisiert?
Wer befasst sich mit Verkehr?

¹⁴ Um Missverständnisse bzw. Fehlzuordnungen bei den Kookkurrenzergebnissen auszuschließen, werden immer auch die KWICs prüfend mit einbezogen.

Entscheidend für das Vorgehen der Lexikografinnen ist, dass sie nicht alle **(theoretisch) möglichen** Mitspieler(gruppen) erfassen, sondern nur die **im Korpus realisierten** und daran ausgerichtet das komplette Set aus Fragen und Antworten zusammenstellen.¹⁵

2.3.2 Typische Verwendungen

Welcher Art die Informationen sind, die Lexikografinnen unter dieser Überschrift erarbeiten, ist in der Online-Ansicht an der betreffenden Stelle in einer aufklappbaren Infobox dokumentiert (siehe <http://www.elexiko.de/>):

Hier sehen Sie, wie das Stichwort in seiner jeweiligen Lesart im Satz eingebettet ist. Die hier genannten Wortgruppen (syntagmatischen Verbindungen) sind alle häufig und in genau der angegebenen Form im *elexiko*-Korpus belegt und können als typisch gelten. Sie veranschaulichen den typischen (nicht zwangsläufig: den mustergültigen) Gebrauch des Stichworts im Satzzusammenhang. So kann man z. B. erkennen, ob das Stichwort typischerweise im Plural, mit bestimmtem bzw. unbestimmtem Artikel, adverbial, passivisch [...] gebraucht wird. Bei manchen Verwendungen ist der Übergang zur mehr oder weniger festen Verbindung schon zu erkennen. (<http://www.elexiko.de/>: zu Einzelheiten siehe P. Storzjohann, Typische Verwendungen, in diesem Band)

Zur Ermittlung der typischen Verwendungen dienen die syntagmatischen Muster aus der Kookkurrenzliste als entscheidende Quelle.

Um eventuelle Missverständnisse auszuschließen sollte ein vergewissernder Klick und Blick in die dazugehörigen KWICs (und ggf. in die Belege) das Ansetzen einer typischen Verwendung stützen. Ein willkürlich ausgewähltes Beispiel mag der Illustration dienen: Die in *elexiko* angesetzte typische Verwendung *Verkehr in beiden Richtungen* findet sich so oder so ähnlich bereits als syntagmatisches Muster im oben abgebildeten Ausschnitt der Kookkurrenzanalyse:

Richtungen	der den <i>Verkehr</i> aus in beide beiden Richtungen
------------	--

Dem liegen folgende (um die Quellensiglen gekürzte) KWICs zu Grunde:

vor. eine Tunnelröhre für einspurigen **Verkehr** in beide Richtungen zu bauen. Kosten: 370 der Hauptverkehrszeiten unter dichtem **Verkehr** aus allen Richtungen auf der Bundesstraße Arbeiten durchgeführt. die den **Verkehr** der Straßenbahnen in beide Richtungen In Bensheim wird der **Verkehr** in beide Richtungen an der Innenstadt

¹⁵ Lassen sich mithilfe der Kollokationsanalyse keine signifikanten Mitspieler ermitteln, weil sich beispielsweise niedrige Trefferzahlen gleichmäßig auf wenige Kookkurrenzpartner verteilen, so treffen die Lexikografinnen eine Auswahl und kommentieren diese entsprechend.

Sie messe den **Verkehr** in beiden Richtungen, die Geschwindigkeit
 Erst am späten Nachmittag rollte der **Verkehr** wieder flüssig in beide Richtungen.
 über genügend Gleise verfügt, konnte der **Verkehr** in beide Richtungen unbehindert weiter
 Ab Oktober 2001 soll der **Verkehr** in beide Richtungen zweispurig fließen.
 in beiden Richtungen für den **Verkehr** wieder geöffnet.
 Der **Verkehr** verläuft in beiden Richtungen auf nur
 „Während dieser Arbeiten wird der **Verkehr** voll in beiden Richtungen fließen.“
 Dessau-Ost und -Süd wird der **Verkehr** in beiden Richtungen wegen Bauarbeiten
 Ende nächsten Jahres soll der **Verkehr** dann zweispurig in beide Richtungen
 gearbeitet werden muß - über und neben dem **Verkehr** in beiden Richtungen.“
 Derzeit wird der **Verkehr** in beide Richtungen auf der bergseitigen
 Der **Verkehr** rollte in beiden Richtungen über den

Daraus wurde das typische Verwendungsmuster schließlich in der angegebenen Form vom Lexikografen ausgewählt.


Insgesamt wurden für *Verkehr* ‚Fortbewegen‘ nach Überprüfung der syntagmatischen Muster und der dazu gehörigen KWICs folgende typischen Verwendungen abgeleitet und nach zunehmender Komplexität angeordnet.¹⁶

öffentlicher Verkehr
 beim Thema Verkehr
 dem Verkehr übergeben
 den Verkehr aufhalten
 den Verkehr behindern
 den Verkehr blockieren
 den Verkehr einstellen
 den Verkehr regeln
 den Verkehr umleiten
 der Verkehr brummt
 im Verkehr ersticken
 in Sachen Verkehr
 in Verkehr setzen
 mehr Verkehr anziehen
 Sicherheit im Verkehr
 Trunkenheit im Verkehr
 Überwachung des Verkehrs
 Umleitung des Verkehrs
 Umleitung des Verkehrs
 vom Verkehr entlasten
 zum Thema Verkehr
 zusätzlichen Verkehr anziehen
 aus dem Verkehr ziehen
 Ausbau des öffentlichen Verkehrs
 Behinderungen für den Verkehr
 den Verkehr aufnehmen (können)
 den Verkehr flüssiger machen
 den Verkehr großräumig umleiten
 den Verkehr lahm legen

[...] behindert den Verkehr
 [...] zieht Verkehr an
 aus dem Verkehr gezogen werden
 den Verkehr in [...] Bahnen lenken
 den Verkehr nicht aufnehmen (können)
 den Verkehr zum Erliegen bringen
 der Verkehr aus Richtung [...]
 der Verkehr in Richtung [...]
 der Verkehr kam zum Erliegen
 der Verkehr musste umgeleitet werden
 der Verkehr wird nicht behindert
 einspurig für den Verkehr freigeben
 jährlich sterben im Verkehr [...]
 in Sachen Tourismus, Verkehr, Energie
 Institut für Verkehr und Tourismus
 nicht zum Verkehr zugelassene/-s/-r [Verkehrsmittel]
 [...] legte den Verkehr lahm
 ist für den gesamten Verkehr gesperrt
 kombinierter Verkehr (auf Straße und Schiene)
 Ministerium für Wirtschaft, Verkehr und Landesentwicklung
 Ministerium für Wirtschaft, Verkehr und Technologie
 mehr Verkehr auf die Schiene bringen
 ohne auf den Verkehr zu achten
 Senatsverwaltung für Bauen, Stadtentwicklung und Verkehr
 wird für den gesamten Verkehr gesperrt

¹⁶ Obgleich in der Regel alle signifikanten erkennbaren Muster angesetzt werden, kann die große Menge im Einzelfall eine Auswahl nach Frequenz bzw. statistischer Signifikanz durch die LexikografInnen nötig machen.

den Verkehr nicht behindern
 der Verkehr fließt einspurig
 der Verkehr staute sich
 der Verkehr wird behindert
 Drehscheibe für den Verkehr
 Finanzierung des öffentlichen
 Verkehrs
 Förderung des öffentlichen Verkehrs
 für den Verkehr erschließen
 für den Verkehr freigeben
 für den Verkehr öffnen
 für den Verkehr sperren
 in (den) Verkehr bringen
 [nicht] zum Verkehr zugelassen
Verkehr in beiden Richtungen
 vom Verkehr abgeschnitten sein
 vom Verkehr entlastet werden
 wird dem Verkehr übergeben

zu Behinderungen für den Verkehr kommen
 [...] ist für den Verkehr geöffnet
 das Bundesministerium für Verkehr, Bau- und
 Wohnungswesen
 die Gewerkschaft Öffentliche Dienste, Transport
 und Verkehr
 das Großressort für Bauen, Verkehr und Stadtent-
 wicklung
 die traditionellen Dienstleistungsbereiche Handel,
 Gastgewerbe und Verkehr
 Verlagerung des Verkehrs von [...] auf [...] 
 Supersenator für Bauen, Wohnen, Verkehr, Stadt-
 entwicklung und Umweltschutz
 die Branchen Energie, Verkehr, Biotechnologie,
 Audiovision, Information und Telekommunikation
 das Eidgenössische Departement für Umwelt, Ver-
 kehr, Energie und Kommunikation

2.3.3 Sinnverwandte Wörter

Welcher Art die Informationen sind, die Lexikografinnen unter dieser Überschrift erarbeiten, ist in der Online-Ansicht an der betreffenden Stelle in einer aufklappbaren Infobox dokumentiert (siehe <http://www.elexiko.de/>):

Sinnverwandte Wörter stehen in Beziehungen, so genannter paradigmatischer Relationen oder Sinnrelationen, wie z. B. Über- und Unterordnung, Bedeutungs-
 gleichheit und Bedeutungsgegensätzlichkeit. Damit wird der Wortschatz sowohl
 horizontal als auch vertikal gegliedert. Diese Sinnrelationen können nicht insge-
 samt für ein Stichwort angegeben werden, da sie immer kontextgebunden sind und
 sich auf einen gemeinsamen Textzusammenhang beziehen. Deshalb sind in *elexiko*
 Synonyme, Hyperonyme, Antonyme etc. für einzelne Lesarten angegeben, deren
 relevante Kontexte durch Beispielbelege veranschaulicht werden. Im Unterschied
 zu anderen Wörterbüchern werden in *elexiko* mehrere Arten von Gegensatzbezie-
 hungen unterschieden [...]. Außerdem ist in *elexiko* jedes sinnverwandte Partner-
 wort als Link realisiert, sodass man direkt zu dem gewünschten Partnerwort sprin-
 gen kann. (<http://www.elexiko.de/>; nähere Einzelheiten dazu siehe P. Storjohann,
 Paradigmatische Relationen, in diesem Band)

Die Vorgehensweise zur Ermittlung paradigmatischer Partnerwörter, insbe-
 sondere die Corpus-Driven-Methode und die Corpus-Based-Methode, sowie
 deren Wechselspiel führt P. Storjohann im Beitrag ‚Paradigmatische Relatio-
 nen‘ am Beispiel *flexibel* ausführlich vor. Daher genügt es, das Verfahren für
Verkehr/‚Fortbewegen‘ kurz zu rekapitulieren.

Neben den als inkompatibel eingestuften Partnern *Energie*, *Finanz*, *Handel*, *Technik*, *Kommunikation*, *Mobilität*, *Transport* und *Umwelt* wurde für *Verkehr* in der Lesart ‚Fortbewegen‘ ein Parteronym ermittelt: *Infrastruktur*.

Dieses Wort taucht bereits als Partner im oben abgedruckten Ausschnitt der Kookkurrenzanalyse auf, was auf ein usuelles Miteinandervorkommen beider Wörter in Texten des *elexiko*-Korpus schließen lässt:

Infrastruktur

die ... Infrastruktur [für] *Verkehr*

Natürlich stehen nicht alle häufig miteinander vorkommenden Wörter auch in sinnrelationaler Beziehung zueinander. Die Lexikografinnen müssen hinter die Zeilen der Kookkurrenzanalyse blicken, das Ergebnis transparent machen; um – ganz konkret – die (kompetenzgestützt vermutete) Sinnrelation der Parteronymie zwischen *Verkehr* und *Infrastruktur* im *elexiko*-Korpus nachweisen (oder widerlegen) zu können, wird man von der Zeile des Analyseergebnisses in die KWIC- und die Belegansicht wechseln, um dort nach Indikatoren pro vs. contra zu suchen. Bei hochfrequenten Stichwörtern wird die Trefferzahl durch die statistische Zufallsauswahl und durch die Standardparametereinstellungen der Kookkurrenzanalyse z. T. erheblich reduziert. Um das Miteinandervorkommen zweier Wörter im Ergebnis der Kookkurrenzanalyse (= CD-Verfahren), mithin einen sinnrelationalen Zusammenhang gezielt und umfassend nochmals im gesamten *elexiko*-Korpus zu überprüfen, bietet sich eine neuerliche Cosmas-II-Abfrage an (= CB-Verfahren). In der Form „&Verkehr /s/ &Infrastruktur“ kann das Vorkommen beider Lemmata im gleichen Satz bzw. im unmittelbar nachfolgenden oder vorangehenden (Abstandsoperator: „/s/“) eruiert werden: Eine Wortform von *Verkehr* kommt mit einer Wortform von *Infrastruktur* im *elexiko*-Korpus insgesamt 618 mal im Kontext von zwei Sätzen vor.

Die Durchsicht der KWICs dieser Abfrage führt zu einer Reihe Zeilen wie diesen:

der Bundesregierung genauso wie die **Infrastruktur** in den Bereichen Energie, **Verkehr** gesetzlich zwingende Bestimmung, dass die **Infrastruktur** für den **Verkehr** in die Höhe oder in

Mit solchen KWICs und ggf. den gesamten dazugehörigen Belegen kann das semantische Verhältnis zwischen *Infrastruktur* und *Verkehr* näher bestimmt werden; hierauf gründet sich letztlich die Entscheidung, *Infrastruktur* als Parteronym anzusetzen, d. h.: *Verkehr* ist im *elexiko*-Korpus inhaltlich ein (Bestand-)teil des Relationspartners *Infrastruktur* (Parteronym) und nicht umgekehrt (siehe dazu P. Storjohann, Paradigmatische Relationen, in diesem Band).

2.3.4 Besonderheiten des Gebrauchs

Welcher Art die Informationen sind, die LexikografInnen unter dieser Überschrift erarbeiten, ist in der Online-Ansicht an der betreffenden Stelle in einer aufklappbaren Infobox dokumentiert (siehe <http://www.elexiko.de/>):

Hier werden Angaben über eine eventuelle Einstellung des Sprechers, den Situationsbezug, das Sachgebiet und eine mögliche Textsortenbindung des Stichworts in der jeweiligen Lesart gemacht. Diese Angaben ersetzen die in vielen Wörterbüchern üblichen Stil-, Sach-, Fachgebiets- und Konnotationsmarkierungen [...]. (<http://www.elexiko.de/>; nähere Einzelheiten dazu siehe U. Haß, Besonderheiten des Gebrauchs, in diesem Band)

Um Gebrauchsbesonderheiten einer Lesart im *elexiko*-Korpus auf die Spur zu kommen, können die Kookkurrenzanalyse bzw. die KWICs allenfalls erste Spuren aufzeigen; diese können von ‚markierten‘ Mitspielern herrühren, die auf Verwendungsspezifika (in Bezug auf Sprechereinstellung, Situation, Domäne, Textsortenspezifika) hindeuten; Aufgabe der LexikografInnen ist es, diesen Spuren v. a. in den entsprechenden Belegen genauer nachzugehen. Gelangen sie hierbei zu dem Schluss, dass die Verwendung einer Lesart an bestimmte Gebrauchsbesonderheiten geknüpft ist, so wird ein entsprechender Vermerk formuliert und mit einem Textausschnitt aus dem *elexiko*-Korpus belegt. *Verkehr* ist in der Lesart ‚Beischlaf‘ in der Regel durch eindeutig situationsbezogene Verwendung gekennzeichnet, was den Lexikografen zu folgendem Hinweis veranlasst hat (vgl. Abbildung 5).

Ein Blick in die Belege macht weiterhin die Notwendigkeit deutlich, für die Verwendung von *Verkehr* in der Lesart ‚Beischlaf‘ auch eine Sachgebietsangabe zu formulieren und zu belegen (vgl. Abbildung 6)¹⁷:

¹⁷ Sachgebietsangaben können sich auch auf die signifikante Verwendung der Lesart innerhalb bestimmter, ausgezeichnete Testsorten bzw. Ressorts im *elexiko*-Korpus gründen.

Verkehr Lesart ‚Beischlaf‘

Besonderheiten des Gebrauchs

Kommentar:

Verkehr wird von SprecherInnen stets dann benutzt, wenn sie den HörerInnen gegenüber deutlich machen wollen, dass sie sich von der jeweils bezeichneten Sache distanzieren und innerlich unbeteiligt sind. Dies geschieht, indem sie für den bezeichneten Vorgang eine möglichst sachliche bzw. nüchterne Bezeichnung wählen.

Beleg:

„Ganz genau informiert der Salzburger Weihbischof Andreas Laun: „Für alle unverheirateten Paare und Homosexuelle gilt: Gegenüber dem ungeschützten **Verkehr** ist der kondom-geschützte das kleinere Übel - aber wohlgemerkt, er bleibt ein Übel, eine Sünde“ (Oberösterreichische Nachrichten, 14.02.1996, Bei uns bleibt der Gummi üble Sünde)

Abbildung 5: Kommentar und Beleg zu Besonderheiten des Gebrauchs im Wortartikel *Verkehr*, Lesart ‚Beischlaf‘

Verkehr Lesart ‚Beischlaf‘

Besonderheiten des Gebrauchs

Verkehr wird im *lexiko*-Korpus besonders oft im institutionellen, offiziellen Sprachgebrauch der Kirche, der Polizei, im Rechtswesen und in Gesetzestexten verwendet, um den Beischlaf zu bezeichnen (vgl. den Beleg).

Beleg:

„Vor dem Geschworenengericht wird - im unmittelbaren Verfahren - der Sachverhalt genauestens aufgearbeitet. Der Sachverhalt, das bedeutet in diesem Fall: Sex. Vor einem Gericht darüber zu sprechen, ist gewiss nicht einfach, weil den meisten Menschen das präzise Vokabular dazu fehlt. Sprachliche Präzision ist aber in einem Prozess absolut erforderlich. Und so kam es gestern zu einigen Missverständnissen. Es begann damit, dass der Angeklagte einige Male die Begriffe „oral“ und „anal“ verwechselte, was seinen Verteidiger zum gereizten Zwischenruf veranlasste, man solle doch nun endlich die umgangssprachlichen deutschen Wörter dafür benutzen. Den Ausdruck „geschlechtlicher **Verkehr**“ interpretierte der Angeklagte wiederum anders als der Staatsanwalt - für Adrian gehörte oraler Sex nicht dazu (Zürcher Tagesanzeiger, 07.07.1999, S. 17, Über Sex sprechen.)

Abbildung 6: Sachgebietsangabe und Beleg im Wortartikel *Verkehr*, Lesart ‚Beischlaf‘

Die Prüfung der Kookkurrenzpartner, KWICs und der Belege machte auch für *Verkehr* ‚Fortbewegen‘ die Formulierung von Angaben zu Besonderheiten des Gebrauchs notwendig (siehe Artikel *Verkehr* unter <http://www.elexiko.de/>).

Entscheidend ist, dass diese in *elexiko* (anders als vielfach in der bisherigen Praxis üblich) stets

- korpusbasiert,
- in Sätzen formuliert
- und durch Belege illustriert

angegeben werden.

2.3.5 Grammatik

Welcher Art die Informationen sind, die LexikografInnen unter dieser Überschrift erarbeiten, ist in der Online-Ansicht an der betreffenden Stelle in einer aufklappbaren Infobox dokumentiert (siehe <http://www.elexiko.de/>):

In *elexiko* werden für alle Stichwörter grammatische Angaben gemacht, wobei diese sich im Einzelfall auf die Angabe der Wortart beschränken können (z. B. bei Junktoren). Meistens werden in *elexiko* aber umfangreichere grammatische Angaben geboten als in traditionellen Wörterbüchern (z. B. werden bei den Verben die entsprechenden Satzbaupläne systematisch aufgelistet) [...].

Alle grammatischen Angaben sind aus dem *elexiko*-Korpus gewonnen; sind einzelne Deklinations- oder Flexionsformen der Substantive, Adjektive, Pronomen und der Verben aus systematischen Gründen ergänzt worden, können Sie dies an der Markierung „(nicht im *elexiko*-Korpus belegt)“ bei der entsprechenden Form erkennen.

Stehen konkurrierende Formen nebeneinander (z. B. Komparativ von *rot*: *röter* / *roter*), erhalten beide Formen eine Markierung ihrer jeweiligen Verwendungshäufigkeit im *elexiko*-Korpus [...]. (<http://www.elexiko.de/>; nähere Einzelheiten dazu siehe A. Klosa, Grammatik, in diesem Band)

Vor allem im Bereich Grammatik können spezielle Wortformabfragen in COSMAS II nötig werden, um gezielt das Vorkommen ungewöhnlicher und/oder seltener Paradigmenformen im *elexiko*-Korpus zu prüfen, damit LexikografInnen die im Rahmen der DTD geforderten Einträge bestimmter grammatischer Paradigmenformen¹⁸ und Verwendungshäufigkeitsangaben machen können.

Verkehre, der Plural von *Verkehr* ist gemäß einer Suchwortanfrage im *elexiko*-Korpus insgesamt nur 392 mal belegt gegenüber insgesamt knapp 87000 Treffern für das Stichwort *Verkehr*. Ein Blick in die betreffenden Belege zeigt,

¹⁸ Z. B. der Plural bei Substantiven, die 1. Pers. Sing. Prät. aktiv bei Verben, Komperativ und Superlativ bei Adjektiven, siehe dazu A. Klosa, Grammatik, in diesem Band.

dass der Plural nicht nur für die als durativen Gattungsprädikator eingestufte Lesart ‚Beischlaf‘ Verwendung findet, sondern ungewöhnlicherweise auch für die als Kollektivum (!) klassifizierte Lesart ‚Fortbewegen‘. Eine mögliche Erklärung hierfür hat der Lexikograf als Kommentar formuliert, belegt und mit einem sprachreflexiven Hinweis (siehe dazu: A. Klosa, Sprachkritik und Sprachreflexion, in diesem Band) versehen (vgl. Abbildung 7).

Verkehr Lesart ‚Fortbewegen‘

Grammatik

Kommentar:

Im *lexiko*-Korpus ist der Plural **Verkehre** selten und in der Regel domänenspezifisch- bzw. fachsprachlich belegt. Die Pluralform **Verkehre** wird oft synonymisch zu **Verkehr** verwendet. Im Gegensatz dazu verdeutlicht ihr Gebrauch jedoch, dass die kollektiv mit **Verkehr** bezeichneten Personen- und Güterfortbewegung auch differenziert (betrachtet) werden kann: Die Sammelbezeichnung umfasst nämlich den Einsatz unterschiedlicher Verkehrsmittel. Außerdem wird mit dem Plural **Verkehre** auf das Verkehren, d. h. das konkrete Fahren, Bezug genommen, sowie auf unterschiedliche Möglichkeiten der Datenübertragung.

Beleg:

„Ja, aber es gibt auch eine ganze Reihe notwendiger Innenstadt- **Verkehre** – Lieferverkehre, Polizei, Krankenwagen usw. Wir sind ja heute in der Situation, daß durch die Verkehrszunahme der letzten Jahre – der große Sprung war noch einmal von 1985 bis 1990 – die Autos sich zunehmend selber blockieren. Deshalb stehen wir heute vor allem vor der Aufgabe, die notwendigen **Verkehre** wieder flüssiger zu machen. Das heißt, der ÖPNV, das Fahrradfahren und auch das Zufußgehen müssen wieder schmackhaft gemacht werden. (die tageszeitung, 29.12.1993, S. 19, „Überzogene Maßnahmen führen zum Gegenteil“.)

Sprachreflexiver Hinweis:

Die fachsprachliche Beschränkung des Plurals **Verkehre** findet sich auch in der sprachpflegerischen Literatur.

Weitere Informationen:

Duden - Richtiges und gutes Deutsch. Zweifelsfälle der deutschen Sprache von A - Z. Duden-Verlag: 5., neu bearbeitete Auflage Mannheim/Leipzig/Wien/Zürich 2000, S. 898.

Abbildung 7: Kommentar, Beleg und Sprachreflexiver Hinweis zur Pluralbildung im Wortartikel *Verkehr*, Lesart ‚Fortbewegen‘

Falls – anders als beim vorliegenden Beispiel – die Analyse der syntagmatischen Muster in der Kookkurrenzliste bzw. der dazugehörigen KWICs¹⁹ vor allem bei Verben, aber ggf. auch bei Substantiven und Adjektiven die Angabe von Valenzen erforderlich macht, werden entsprechende Phrasenmusterangaben selbstverständlich davon abgeleitet und entsprechend abstrahiert formuliert (siehe A. Klosa, Grammatik, in diesem Band, 291).

Mit dem Eintragen der Informationen zum Schwerpunkt Grammatik ist ein ‚Durchgang‘ bei der Wortartikelproduktion abgeschlossen. Besitzt ein Stichwort mehr als eine Lesart, wenden sich die Lexikografen nun dieser bzw. diesen zu, um sie analog zum skizzierten Vorgehen zu *lexikografisch* zu erfassen.²⁰ Der komplette Wortartikel wird schließlich – Validität vorausgesetzt (siehe C. Müller-Spitzer, Modellierung, in diesem Band, 56 f.) – in der Datenbank abgelegt und dort für die Online-Präsentation freigegeben.

2.4 Ergänzende Informationen zu den lesartenübergreifenden bzw. lesartenbezogenen Angaben)

Die vorgestellten, von den LexikografInnen obligatorisch zu ermittelnden Daten zu den einzelnen Informationsschwerpunkten werden ergänzt durch Angabetypen, die entweder schwerpunktübergreifend angesetzt werden können, oder die zur Zeit lediglich fakultativ auszufüllen sind. Beispiele für Erstere wurden oben bereits erwähnt; sie lassen sich unter der Überschrift Sprachkritik und Sprachreflexion zusammenfassen (siehe A. Klosa, Sprachkritik und Sprachreflexion, in diesem Band), aber auch der Einsatz von Belegen (siehe A. Klosa, Belege, in diesem Band) kann hier angeführt werden. Ein Beispiel für Zweiteres sind Informationen zur Diachronie in *lexiko*, sie „sollen zunächst nur geschichts- oder diskursrelevanten Schlüsselwörtern vorbehalten sein“ (P. Storzjohann, Diachrone Angaben, in diesem Band, 316).

Eine Mittelstellung nehmen Angaben zur nationalen Verteilung des Stichwortes ein. Sie setzen zweierlei voraus: Die Lesartendisambiguierung sowie eine Korpusbefragung, deren Ergebnis die Liste aller Quellen enthält, in denen das Stichwort vorkommt. Indem die LexikografInnen durch einfachen Maus-

¹⁹ Hierbei bietet sich die alphabetische Sortierung der KWICs an, gestaffelt nach drei Kriterien: 1. Suchwort, 2. erster Nachfolger und 3. zweiter Nachfolger. Auf diese Weise lassen sich z. B. präpositionale Anschlüsse leichter auffinden

²⁰ Bei polysemen Stichwörtern bestimmt die gerade zur Bearbeitung anstehende Lesart zwar die Auswertung der Korpusdaten. Die Untersuchung von Kookkurrenzlisten aus dem *lexiko*-Korpus ist jedoch ökonomischer, wenn man dabei auch stets die anderen Bedeutungen des Stichworts im Hinterkopf hat und eventuelle Auffälligkeiten in einem Zuge markiert bzw. notiert.

klick von der ‚Quellen-‘ in die ‚KWIC-‘ bzw. ‚Belegansicht‘ wechseln, können sie feststellen, ob das Stichwort (in all seinen Lesarten) bzw. die einzelnen unterschiedenen Lesarten in bundesdeutschen, österreichischen, schweizerischen und/oder aus der DDR stammenden Korpus-texten vorkommen (zur diesbezüglichen Zusammensetzung des *ellexiko*-Korpus siehe P. Storzjohann, Das *ellexiko*-Korpus, in diesem Band). Wenn keine lesartenspezifischen Unterschiede vorliegen, werden Angaben zur nationalen Verteilung stets lesartenübergreifend eingetragen. Andernfalls, und dies trifft auf *Verkehr* zu, erfolgen lesartenspezifisch differenzierte Einträge (siehe Wortartikel *Verkehr* unter <http://www.ellexiko.de/>).

Obleich das Prinzip der Korpusbasiertheit die lexikografische Arbeit dominiert, ist selbstverständlich auch in *ellexiko* das Zurateziehen von Nachschlagewerken ein integrativer Bestandteil im Arbeitsprozess der LexikografInnen, beispielsweise um sinnverwandte Wörter gegenzuprüfen, um sich Rat in Normfragen zu holen, oder um etymologischen Fragestellungen nachzugehen.

3. *ellexikografie* – Ausblick

Ein Überblick über den lexikografischen Arbeitsplatz wäre sicherlich unvollständig, würde man nicht abschließend noch einige Perspektiven und Desiderate ansprechen. Die Arbeit der ‚*ellexikografInnen*‘ erschöpft sich im Augenblick nicht in der zweifelsfrei anspruchsvollen Produktion von Wortartikeln; dieser für ein lexikografisches Projekt natürlich äußerst wichtige Prozess ist nämlich nur durchführbar, wenn gleichzeitig eine geeignete Infrastruktur bereitgestellt und weiterentwickelt wird. Obgleich – wie die Beiträge dieses Bandes zeigen – sehr viel geleistet wurde, befindet sich *ellexiko* als Ganzes immer noch in unterschiedlichen Stadien der Realisierungsphase. Als Schwerpunkte für den zukünftigen Ausbau von *ellexiko* sind im Augenblick u. a. geplant:

1. Ausbau des Datenbestandes und des Informationspotentials

Eine Besonderheit von *ellexiko* rührt von der elektronischen Publikationsform her: Die LexikografInnen haben nicht mehr alphabetisch definierte Buchstabenstrecken abzuarbeiten, um schließlich den Gesamtwortschatz von A – Z zu erschließen. Vielmehr bestimmt die Auswahl inhaltlicher bzw. thematisch definierter Wortschatzbereiche (sog. Module) den Arbeitsfortgang.²¹

²¹ Nach dem abgeschlossenen *ellexiko*-Kernmodul ‚Demonstrationswortschatz‘ ist als nächstes geplant, ein ‚Lexikon zum öffentlichen Sprachgebrauch‘ zu erarbeiten. Dieser Arbeitstitel liegt nahe, wenn man die in Frage kommende Menge eigentlich mittels Korpusfrequenz aus-

Wortartikel werden zudem nicht nur von Lexikografinnen mit angemessener Detailtiefe, gewissermaßen vertikal in das System eingespeist, sondern es ist geplant, verstärkt auf Breite angelegte Datenmengen gewissermaßen horizontal über die gesamte Stichwortmenge hinweg automatisch zu erzeugen und in die Wortartikel zu integrieren.

Obschon die Konzeption für Angaben zur Wortbildungsproduktivität (siehe A. Klosa, Wortbildung, in diesem Band) und Diachronie (siehe P. Storjohann, Diachrone Angaben, in diesem Band) als auch deren Datenmodellierung bereits vorliegen, wurden bislang aus Zeit- und Kapazitätsgründen dazu keine Daten erarbeitet; sie ‚fehlen‘ also auch im Demonstrationswortschatz.

2. Recherchierbarkeit und Präsentation der Daten

Den Lexikografinnen ist natürlich sehr daran gelegen, dass die in XML mittels DTD erfassten und ausgezeichneten Daten der *lexiko*-Wortartikel online recherchiert bzw. den NutzerInnen entsprechend präsentiert werden. Was Letzteres betrifft, so sei nochmals darauf hingewiesen, dass je nach Nutzertyp und Benutzungssituation unterschiedlichste Präsentationsformen auf der gleichen Datenbasis möglich sind – wenn man die hierfür notwendigen Stylesheets programmiert.²² Voraussetzung ist natürlich die Präzisierung der Adressaten(gruppen) sowie der Nutzungssituationen. Neben der Arbeit am Ausbau von *lexiko* sind intensive Anstrengungen nötig, um die in der Datenbank abgelegten und ausgezeichneten Informationen auch in jeweils adäquater und ansprechender Weise auf die Bildschirme unterschiedlichster NutzerInnen mit unterschiedlichsten Recherchebedürfnissen zu bringen.

3. Dokumentation

Ein weiterer Baustein ist die Vollendung des Redaktionshandbuchs und dessen anschließende Integration in die bereits intern vorliegende DTD-Dokumentation (siehe C. Müller-Spitzer, Modellierung, in diesem Band, 60). Ein entsprechend ausgearbeiteter Leitfaden ist auch notwendig, um zukünftige ‚Lexikografinnen‘ mit dem *lexiko*-Arbeitsplatz zügig vertraut zu machen.

gewählter *lexiko*-Stichwörter auch unter thematisch-diskursiven Aspekten analysiert und einzuordnen versucht.

²²

Dabei ist u.a. zu entscheiden, welche Daten wie (formatiert) jeweils angezeigt werden, ob und in welcher Sprache (für Laien oder Experten) sie kommentiert werden sollen usw.

Auf ein Desiderat, dessen Einlösung mit zunehmendem Bestand an fertigen *elexiko*-Wortartikeln dringlicher wird, sei abschließend noch hingewiesen: Es ist die komfortable Verwaltung der Wortartikelinstanzen im Rahmen eines Redaktionssystems; damit lassen sich etwa Autorenrecherchen vereinfachen, etwa mit dem Ziel, Überarbeitungen zu erleichtern, bestimmte Problemlösungsstrategien in älteren Wortartikeln nachzuprüfen, Bestände von Wortartikel zu homogenisieren usw.; auch die Regelung von Zugriffsrechten oder die Durchführung von Sicherungsroutinen könnte auf diesem Wege in Zukunft effizienter, der lexikografische Arbeitsplatz noch komfortabler gestaltet werden.

4. Literatur

4.1 Forschungsliteratur

Belica, Cyril (1995): Statistische Kollokationsanalyse und Clustering. COSMAS-Korpusanalysemodul. Mannheim.

4.2 Internetressourcen

elexiko: <http://www.elexiko.de/> (letzter Zugang: September 2004)

COSMAS II: <http://www.ids-mannheim.de/cosmas2/> (letzter Zugang: September 2004)

Statistische Kollokationsanalyse und Clustering: <http://www.ids-mannheim.de/kt/projekte/methoden/> (letzter Zugang: September 2004)